On the Scalability of Duty-Cycled LoRa Networks with Imperfect SF Orthogonality

Bouazizi, Y., Benkhelifa, F., ElSawy, H. & McCann, J.

DOI 10.1109/lwc.2022.3200912
ISSN 2162-2337
ESSN 2162-2345

Publisher: IEEE

On the Scalability of Duty-Cycled LoRa Networks with Imperfect SF Orthogonality

Yathreb Bouazizi, Graduate Student Member, IEEE, Fatma Benkhelifa, Member, IEEE, Hesham ElSawy, Senior Member, IEEE, and Julie A. McCann, Member, IEEE

***Abstract*—This papers uses stochastic geometry and queuing theory to study the scalability of long-range (LoRa) networks, accounting for duty cycling restrictions and imperfect spreading factor (SFs) orthogonality. The scalability is characterised by the joint boundaries of device density and traffic intensity per device. Novel cross-correlation factors are used to quantify imperfect SF-orthogonality. Our results show that a proper characterisation of LoRa orthogonality extends the scalability of the network. They also highlight that for low/medium densities decreasing the SF extends the spanned spectrum of sensing applications characterised by their traffic requirements (i.e. sensing rate). However, for high density ($> 10^4$ nodes/Km^2), the Pareto frontiers converge to a stability limit governed by the SF allocation scheme and the predefined capture thresholds. The results further evince the importance of capturing threshold distribution among the SFs to mitigate the unfair latency.**

***Index Terms*—LoRa, SF-Allocation, Stochastic Geometry, Queuing Theory, Coverage Probability, Stability Analysis.**

I. Introduction

Long-Range (LoRa) is a license-exempt low power wide area (LPWA) technology that has attracted the attention of commercial stakeholders due to its low-cost-low-energy advantage.cThe academic community has also become increasingly interested in LoRa as its efficiency maximization presents many challenges. The predicted massive deployment of Internet of Things (IoT) devices raises scalability concerns. Therefore understanding LoRa's limits in terms of device density and bounds on traffic requirements per device is fundamental. To address such concerns, several models have been developed to mathematically characterize LoRa's scalability [1]–[4]. However, they are primarily focused on spatial coverage analysis and overlook the temporal component of the traffic, assuming a saturated buffer. Such assumption ignores the practicalities of the temporal evolution of packet generation and accumulation in devices buffers. Traffic temporal aspects are typically described using discrete-time Markov chains (DTMCs). Such DTMCs are used to characterize latency in LoRa networks and to develop optimized transmission policies for the nodes [5], [6]. However, the analysis in [5], [6] is based on simplified collision models that neglect spatial traffic behaviours (e.g., network geometry and mutual interference) and limit the network size to a few nodes, which is not practical for applications with massive ubiquitous connectivity. The work in [5] analyzes the Medium Access (MAC)-layer performance of LoRa devices. Assuming perfect-orthogonality[1], the network is modelled as the aggregation of independent sub-networks.Each sub-network operates on a different spreading factor (SF), which fails to capture the actual interaction between SFs that are, in reality, not perfectly orthogonal [7]–[10]; and makes the analysis optimistic. Indeed, as highlighted in [9] SFs' imperfect orthogonality has non-negligible impact on LoRa throughput. The authors in [6] propose a DTMC model that accounts for LoRa Duty Cycle (DC) restrictions. However, by following a node-centric approach, they ignore interference, making their model oblivious to LoRa physical layer attributes (e.g., channel, distance, interference). Their restrictive assumption that no queuing events are allowed at the node level implies that events happening during the node's DC state will be discarded, resulting in data loss.

To the best of the authors' knowledge, this paper is the first to bridge the gap between stand-alone queuing approaches and scalability-oriented analysis for LoRa networks. We propose a novel inclusive spatiotemporal model where queuing theory (QT) and stochastic geometry (SG) are combined via a 2D DTMC that tracks the node's data buffer and protocol state. A more realistic mutual interference scenario is adopted by accounting for SF-imperfect orthogonality. From the temporal perspective, practical DC restrictions are incorporated into the DTMC. Under different SF-orthogonality assumptions, the per-SF scalability-stability trade-off is characterized using Pareto frontiers. Further, the impact of network parameters on the Per-SF latency is studied.

A. Notation & Organization

We use lowercase/uppercase bold math fonts to denote vectors/ matrices. $\mathcal{L}_X/\mathbb{E}_X$ denote the Laplace/expectation of X. The probability complement is denoted as $\overline{(.)}$. The paper is organized as follows; Section II details the system model, while Section III is dedicated to the analytical framework.Numerical results and the conclusion are presented in sections IV and V.

Y. Bouazizi and J.A. McCann are with the Computing department, Imperial College London, London, United Kingdom (e-mail: y.bouazizi18,j.mccann}@imperial.ac.uk).
F. Benkhelifa is with the School of Computing, Electronics and Mathematics, Coventry University, Coventry, United Kingdom (e-mail: ad8904@coventry.ac.uk).
H. ElSawy is with the School of Computing, Queen's University, Kingston, ON, Canada (email:hesham.elsawy@queensu.ca)

This work is supported by the Singapore Ministry of National Development and the National Research Foundation, Prime Minister's Office under the Land and Liveability National Innovation Challenge (L2NIC) Research Programme (L2 NIC Award No. L2NICTDF1-2017-3). Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not reflect the views of the Singapore Ministry of National Development and National Research Foundation.

[1]Perfect-orthogonality assumes that only same spreading factors interfere.

II. System Model

A. Spatial Distribution and Propagation model

We consider a single-cell[2] LoRa network where nodes are randomly scattered around the gateway according to a homogeneous Poisson Point Process (PPP) Φ_L of intensity λ_L. Each node is assigned a given SF_q and shares the same bandwidth (BW) to communicate to a central gateway. To account for the traffic requirements and DC states, let $\tilde{\Phi}_L$ be the thinned process of concurrently transmitting nodes. The subset of active nodes using SF_q is $\tilde{\Phi}_{L,q}$. We denote $\mathbf{\Delta} = [\delta_1, \ldots, \delta_q, \ldots, \delta_N]$ the activity vector of different SFs.

In line with the characteristics of low-mobility smart sensing applications, the realization of the network is random yet static, while the fading realization and data generation vary with time. We consider a power-law path-loss propagation model where the signal attenuates with the propagation distance at the rate $\alpha r^{-\eta}$, $\eta > 2$ is the path-loss exponent, and α is the frequency dependant factor. Further, we have Rayleigh block fading channels with unit-mean exponentially distributed gains i.e. $g_i \sim exp(1)$. For tractability reasons, channel gains are assumed to be independent of space and time dimensions.

B. Channel Access and Queuing Model

We assume a time-slotted system. All LoRa nodes sense for the same application characterized by packet size L and per-slot data arrival rate $a \in [0, 1]$ linked to the sensing frequency. At each device, packets are generated according to a geometric distribution with parameter a. The generated packets are stored in a First In First Out (FIFO) queue. The head of the line packet is transmitted at the beginning of the slot. As this paper's objective is to explore LoRa uplink transmissions, we assume error-free DL communications.

C. LoRa Physical Parameters and Operations

LoRa relies on an adapted chirp spread spectrum modulation scheme. A LoRa device transmitting at SF_q has 2^{SF_q} symbols $\{0, ..., 2^{SF_q} - 1\}$ each of SF_q bits. There are two main assumptions in the literature about the interaction between LoRa SFs: "Perfect orthogonality" that only accounts for the interference from nodes using the same SF [2], [5], [6] and "Imperfect orthogonality" [1], [3], [4], [9], [10] all SFs interfere. While the former assumption overestimates the performance, the latter is a worst-case scenario that does not account for the different SFs' impact on the mutual interference magnitude. To mitigate this, [1], [3] used partial temporal overlap for power equalisation purposes in the SINR formulations. Our study examines a more realistic and practically more useful interference scenario of imperfect orthogonality. We use the orthogonality coefficients β_{q_1,q_2} summarized in Table I and obtained using cross-correlation functions analysis. For more details about their derivations and verification, please refer to [8], which comprehensively studies LoRa modulation.

Let N be the total number of available SFs ranging from SF_1 to SF_N. Since LoRa symbol duration is expressed as $t_{s,q} = \frac{2^{SF_q}}{BW}$, l_q the time on-air (ToA) of packet transmission using SF_q, doubles from SF_q to SF_{q+1}. As a license-free technology, LoRa is subject to restrictions on access to the shared medium. These restrictions vary between the bands and can be based on DC regimes. LoRa DC can be 1% or 10%. We fix DC to 1%[3] where each transmission during l_q is followed by a silent duration $99 \times l_q$. The adopted system slot is $T_1 = 100 \times l_1$. This choice implies that each node transmits no more than once per slot $\forall$ SF_q [11]. As LoRa DC duration is a function of the ToA, each transmission using SF_q will be followed by n_q DC logical sub-states, $DC_q = \{DC_{q,1}, ..., DC_{q,n_q}\}$, with $n_q = 2^{q-1} - 1$ for $q > 1$ $n_1 = 0$. The SFs are allocated to the nodes following a distance-dependant scheme[4] where each SF occupies a restricted region of width $[\, d_{q-1} - d_q \,]$ (e.g., Equal-Interval-Based (EIB), Equal-Area-Based (EAB) [1], [3]). Each SF experiences interference from the concurrently transmitting SFs (CoSF+Inter-SF) with an orthogonality coefficient β_{q_1,q_2}.

β_{q_1,q_2}	7	8	9	10	11	12
7	1	0.104	0.062	0.041	0.029	0.021
8	0.104	1	0.073	0.043	0.029	0.020
9	0.062	0.073	1	0.052	0.030	0.020
10	0.041	0.043	0.052	1	0.037	0.021
11	0.029	0.029	0.030	0.037	1	0.026
12	0.021	0.020	0.020	0.021	0.026	1

Tab. I: LoRa Spreading Factor Orthogonality Coefficients

III. Spatiotemporal Framework

The proposed framework relies on integrated SG and QT models. SG allows us to derive the coverage probability considering the nodes' spatial distribution while QT tracks the evolution of the nodes' MAC state and determines the nodes' activities. These activities control the interference intensity needed for the coverage probability. The following subsections detail the SG and QT frameworks as well as the iterative solution to solve the interdependence between them.

A. Macroscopic Coverage Analysis with Stochastic Geometry

The received signal to interference and noise ratio (SINR) at the gateway from the desired LoRa node, located at $r_0 = \|x_0\|$ and using SF_{q_0} is given by:

$$\mathrm{SINR}(r_0, q_0) = \frac{\mathcal{P} g_0 \alpha r_0^{-\eta}}{\mathcal{I} + \sigma^2}, \tag{1}$$

where $\mathcal{P}$ the transmit power, $\mathcal{I}$ is the aggregate interference , and σ^2 is the variance of the additive white Gaussian noise (AWGN). In the dBm scale, σ^2 is given by $\sigma^2 = -174 + NF + 10\log(BW)$, where NF is the noise figure of LoRa receivers and BW is the bandwidth. $\mathcal{I}$ is expressed as:

$$\mathcal{I} = \sum_{q \in Q} \sum_{x_j \in \tilde{\Phi}_{L,q} \setminus x_0} \mathcal{P} \alpha \|x_j\|^{-\eta} \beta_{q_0,q} g_j, \tag{2}$$

A LoRa gateway successfully decodes a packet if it has a free decoding path [10][5] and the instantaneous SINR surpasses a

[2]While multi-cell topologies could better depict dense LoRa deployments especially in mobility scenarios, single-cell LoRa systems still have their significance for deployments in small towns and for naturally-clustered sensing applications such as building, industrial, and agricultural monitoring.

[3]The scalability of (Listen before talk) LBT-based networks available for LoRa frequency plan KR920 is deferred for a future work.

[4]The distance-based division of the cell justifies the assumption that devices within each SF_q-region exhibit i.i.d steady state queue distribution,

[5]In this study, we assume unsaturated demodulating paths. The assignment policies of demodulating paths and the impact of their saturation are out the this paper's scope and will be explored in a future work.

capture threshold γ_q. We denote $\Gamma = [\gamma_1, \gamma_2 ... \gamma_N]$.

$$p_s(r_0, q_0, \gamma_{q_0}) = \mathbb{P}\{\text{SINR}(r_0, q_0) \geq \gamma_{q_0}\} = e^{-\rho_{q_0}\sigma^2}\mathcal{L}_{\mathcal{I}}(\rho_{q_0}), \quad (3)$$

with $\mathbb{P}\{\cdot\}$ being the probability operator and $\rho_{q_0} = \frac{\gamma_{q_0} r_0^{\eta}}{\mathcal{P}\alpha}$.

Theorem 1. *Considering a distance-based SF allocation, the coverage probability, defined as the spatially averaged transmission success probability of SF_{q_0}, is expressed as:*

$$p_{q_0} = \int_{d_{q_0-1}}^{d_{q_0}} p_s(r_0, q_0, \gamma_{q_0}) \frac{2r_0}{d_{q_0}^2 - d_{q_0-1}^2} dr_0, \quad (4)$$

$$p_s(r_0, q_0, \gamma_{q_0}) = e^{-\rho_{q_0}\sigma^2} \prod_{q=1}^{N} e^{-2\pi\delta_q\lambda_L\left(\mathcal{J}(d_q) - \mathcal{J}(d_{q-1})\right)}, \quad (5)$$

$\mathcal{J}(x) = \frac{x^2}{2}\,{}_2F_1\left(1, \frac{2}{\eta}; 1 + \frac{2}{\eta}; -\frac{x^{\eta}}{b}\right)$, $b = \alpha\rho_{q_0}\mathcal{P}\beta_{q_0,q}$, and ${}_2F_1(\cdot)$ the Gaussian hypergeometric function [12].

Proof. The proof is in Appendix A. □

Remark 1. *The study in this paper can be extended to the multi-cell scenario by recalling the stochastic geometry model proposed in [1] where a Matern Cluster Process (MCP) is adopted to depict LoRa multi-gateway topologies.*

B. Microscopic Device-Level Queuing Theory Analysis

Since a single packet can arrive and/or depart from a buffer at a given time slot, each device is characterised by a quasi-birth death (QBD) process [13]. The transition matrix of the DTMC tracks the number of packets at the device buffer and is given, for each SF_q, by :

$$\boldsymbol{P}_q = \begin{pmatrix} \boldsymbol{B}_q & \boldsymbol{C}_q & \mathbf{0} & \mathbf{0} & \mathbf{0} & \dots \\ \boldsymbol{E}_q & \boldsymbol{A_{q,1}} & \boldsymbol{A_{q,0}} & \mathbf{0} & \mathbf{0} & \ddots \\ \mathbf{0} & \boldsymbol{A_{q,2}} & \boldsymbol{A_{q,1}} & \boldsymbol{A_{q,0}} & \mathbf{0} & \ddots \\ \vdots & \ddots & \ddots & \ddots & \ddots & \ddots \end{pmatrix}, \quad (6)$$

where $\boldsymbol{A}_{q,0}$, $\boldsymbol{A}_{q,1}$ and $\boldsymbol{A}_{q,2}$ track the buffer when it increases by one level, stays in the same level, and decreases by one level, respectively. The matrices $\boldsymbol{B}_q$, $\boldsymbol{C}_q$, and $\boldsymbol{E}_q$ track the transition within level 0, from level 0 up to level 1, and from level 1 down to level 0, respectively. Following the Matrix Analytic Method (MAM) [13], we construct the $\boldsymbol{P}_q$. Each device is perceived as a Geo/PH/1 queue represented by the DTMC in Fig.1. The PH-type parameters are the $(n_q + 1) \times (n_q + 1)$ matrix $\boldsymbol{F}_q$ and the $1 \times (n_q + 1)$ initialisation vector $\boldsymbol{\beta}_q = [0, 1, 0, \cdots, 0]$. The transitions at level 0 are captured by $\boldsymbol{M}_{q,0}$. Let $\boldsymbol{s}_q = \boldsymbol{e} - \boldsymbol{F}_q\boldsymbol{e}$.

$$\boldsymbol{F}_q = \begin{bmatrix} 0 & \bar{p}_q & 0 & \dots & 0 \\ 0 & 0 & 1 & \dots & 0 \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ 0 & \dots & 0 & 0 & 1 \\ 1 & \dots & 0 & 0 & 0 \end{bmatrix}, \boldsymbol{M}_{q,0} = \begin{bmatrix} 1 & 0 & 0 & \dots & 0 \\ 0 & 0 & 1 & \dots & 0 \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ 0 & \dots & 0 & 0 & 1 \\ 1 & 0 & \dots & 0 & 0 \end{bmatrix}. \quad (7)$$

As such, the matrices forming $\boldsymbol{P}_q$ have the following expressions: $\boldsymbol{A}_{q,1} = a\boldsymbol{s}_q\boldsymbol{\beta}_q + \bar{a}\boldsymbol{F}_q$, $\boldsymbol{A}_{q,2} = \bar{a}\boldsymbol{s}_q\boldsymbol{\beta}_q$, $\boldsymbol{A}_{q,0} = a\boldsymbol{F}_q$, $\boldsymbol{B}_q = \bar{a}\boldsymbol{M}_{q,0}$, $\boldsymbol{C}_q = a\boldsymbol{M}_{q,0}$, and $\boldsymbol{E}_q = \boldsymbol{A}_{q,2}$. Before analysing the steady state distribution, we check the conditions that ensure stable queues. A stable queue needs to satisfy a departure rate that is higher than the packet arrival rate [13]. The stability check is hence conducted by studying the stationary distribution vector $\boldsymbol{\pi}_q$ of the matrix $\mathbf{A}_q = \mathbf{A_{q,0}} + \mathbf{A_{q,1}} + \mathbf{A_{q,2}}$.

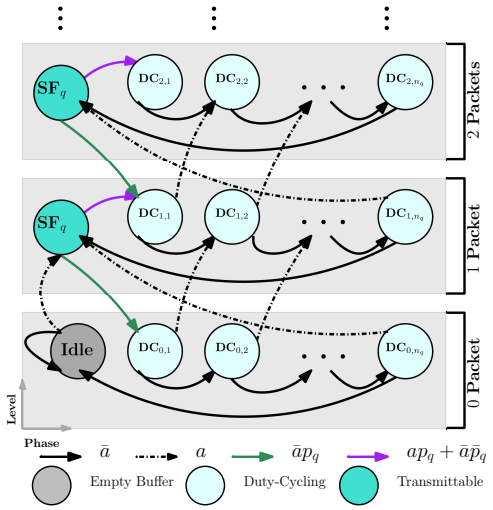


Fig. 1: DTMC of a LoRa node transmitting with SF_q

$\boldsymbol{\pi}_q$ is the unique solution of the equation $\boldsymbol{\pi}_q = \boldsymbol{\pi}_q\mathbf{A}$ with the normalisation condition $\boldsymbol{\pi}_q\mathbf{e} = 1$.

Lemma 1. *For a data buffer to be stable, it needs to satisfy: $\boldsymbol{\pi}_q\mathbf{A_{q,2}}\mathbf{e} > \boldsymbol{\pi}_q\mathbf{A_{q,0}}\mathbf{e}$ where $\boldsymbol{\pi}_q = (\boldsymbol{s}_q\boldsymbol{\beta}_q + \boldsymbol{F}_q - \mathbf{I} + \mathbf{e}\mathbf{e}^T)^{-1}\mathbf{e}$.*

Proof. Follows from Solving the system of equations $\boldsymbol{\pi}_q = \boldsymbol{\pi}_q\mathbf{A}$ and $\boldsymbol{\pi}_q\mathbf{e} = 1$ by recalling [Lemma 1, [14]] and using $\boldsymbol{A_{q,2}}$, $\boldsymbol{A_{q,1}}$, and $\boldsymbol{A_{q,0}}$. □

We define the invariant probability vector with regard to $\boldsymbol{P}_q$ as $\mathbf{x}_q = \begin{bmatrix}\mathbf{x}_{q,0} & \dots & \mathbf{x}_{q,i} \dots\end{bmatrix}$ such that $\mathbf{x}_{q,i} = \begin{bmatrix}x_{q,i,0} & \dots & x_{q,i,n_q+1}\end{bmatrix}$, where $x_{q,i,m}$ is the probability of having i packets in the buffer while being at phase m of the DTMC for SF_q. We obtain $\mathbf{x}_q$ by solving the system of equations defined by $\mathbf{x}_q = \mathbf{x}_q\boldsymbol{P}_q$ and the normalization condition $\mathbf{x}_q\mathbf{e} = 1$.

Theorem 2. *The steady state probabilities $\boldsymbol{x}_q$ of a LoRa node using SF_q is given by:*

$$\boldsymbol{x}_{q,i} = \begin{cases} \boldsymbol{x}_{q,0}\boldsymbol{C}_q\left(\boldsymbol{I} - \boldsymbol{A}_{q,1} - \boldsymbol{R}_q\boldsymbol{A}_{q,2}\right)^{-1} \times \boldsymbol{E}_q\left(\boldsymbol{I} - \boldsymbol{B}_q\right)^{-1}, & \text{if } i = 0, \\ \boldsymbol{x}_{q,0}\boldsymbol{C}_q\left(\boldsymbol{I} - \boldsymbol{A}_{q,1} - \boldsymbol{R}_q\boldsymbol{A}_{q,2}\right)^{-1}, & \text{if } i = 1, \\ \boldsymbol{x}_{q,1}\boldsymbol{R}_q^{i-1}, & \text{if } i > 1, \end{cases} \quad (8)$$

with the normalisation condition $\boldsymbol{x}_{q,0}\boldsymbol{e} + \boldsymbol{x}_{q,1}\left(\boldsymbol{I} - \boldsymbol{R}_q\right)^{-1}\boldsymbol{e} = 1$ being satisfied, and $\boldsymbol{R}_q$ is the MAM rate matrix expressed as:

$$\boldsymbol{R}_q = a\boldsymbol{F}_q(\boldsymbol{I} - a\boldsymbol{s}_q\boldsymbol{\beta}_q - \bar{a}\boldsymbol{F}_q - a\boldsymbol{F}_q\boldsymbol{e}\boldsymbol{\beta}_q)^{-1}. \quad (9)$$

Proof. Following [13, section 5.8], we solve the following system of equations:

$$\begin{pmatrix}\boldsymbol{x}_{q,0} & \boldsymbol{x}_{q,1}\end{pmatrix} = \begin{pmatrix}\boldsymbol{x}_{q,0} & \boldsymbol{x}_{q,1}\end{pmatrix}\begin{pmatrix}\boldsymbol{B}_q & \boldsymbol{C}_q \\ \boldsymbol{E}_q & \boldsymbol{A}_{q,1} + \boldsymbol{R}_q\boldsymbol{A}_{q,2}\end{pmatrix},$$

Using $\boldsymbol{x}_{q,i} = \boldsymbol{x}_{q,i-1}\boldsymbol{R}_q$, the steady state probabilities of the upper levels is deduced. $\boldsymbol{R}_q$ is the minimum non-negative solution of the quadratic equation: $\boldsymbol{R}_q = \boldsymbol{A}_{q,0} + \boldsymbol{R}_q\boldsymbol{A}_{q,1} + \boldsymbol{R}_q^2\boldsymbol{A}_{q,2}$. Since $\boldsymbol{A}_{q,2}$ is of rank 1 [13], $\boldsymbol{R}_q$ is explicitly obtained as in (9). □

Let $\boldsymbol{\varphi}_q = [\varphi_{q,1}, ...\varphi_{q,(n_q+1)}]$ be the marginal distribution of the phases with nonempty data buffers where $\varphi_{q,m}$, corresponds to transmittable or DC states of SF_q.

Corollary 1. *The activity probability of SF_q is given by $\delta_q = \varphi_{q,1}$, with $\boldsymbol{\varphi}_q$ a sub-stochastic vector written as:*

$$\boldsymbol{\varphi}_q = \sum_{i\geq 1}\boldsymbol{x}_{q,i} = \boldsymbol{x}_{q,1}\left(\boldsymbol{I} - \boldsymbol{R}_q\right)^{-1}. \quad (10)$$

Proof. When Lemma 2 is satisfied, $\boldsymbol{R}_q$ has a spectral radius less than 1. Hence, the geometric series generated by $\boldsymbol{R}_q$ converges and the proof follows. □

Remark 2. *If Lemma 1 is not satisfied, the probability to have empty buffer is nullified. Hence, $\boldsymbol{\varphi}_q = \boldsymbol{\pi}_q$ and $\delta_q = \pi_{q,1}$.*

Remark 3. *For the case $q = 1$, the Geo/PH/1 system in* (6) *boils down to a Geo/Geo/1 with:*

$$\boldsymbol{P}_1 = \begin{pmatrix} \bar{a} & a & & & \\ \bar{a}p_1 & \bar{a}\bar{p}_1 + ap_1 & a\bar{p}_1 & & \\ & \bar{a}p_1 & \bar{a}\bar{p}_1 + ap_1 & a\bar{p}_1 & \\ & & \ddots & \ddots & \ddots \end{pmatrix}. \tag{11}$$

If $p_1 > a$, $\boldsymbol{P}_1$ is stable and $\boldsymbol{x}_1$ is given by:

$$x_{1,i} = \begin{cases} \frac{p_1 - a}{p_1}, & \text{if } i = 0, \\ \left(\frac{\bar{p}_1 a}{p_1 \bar{a}}\right)^i \frac{x_0}{\bar{p}_1}, & \text{if } i \geq 1. \end{cases} \tag{12}$$

The activity probability is $\delta_1 = \bar{x}_{1,0}$.

C. Iterative Algorithm and KPIs

The coverage probability p_q in (5) requires different δ_q obtained from the DTMC steady state. On the other side, $\overline{p}_q$ is used in $\boldsymbol{F}_q$. To resolve such interdependence, an iterative solution detailed in Algorithm 1 is proposed.

The Pareto stability frontiers are characterised by the set of parameters $\mathcal{S}_q = \{\{\lambda_L, \gamma_q, a\} \in \mathbb{R}^2 \times [0,1], \boldsymbol{\pi}_q \mathbf{A}_{\mathbf{q},2}\mathbf{e} > \boldsymbol{\pi}_q \mathbf{A}_{\mathbf{q},0}\mathbf{e}\}$. Operating within this region ensures a service rate higher than the arrival rate and guarantees finite packet delays.

We denote W_q the duration of a randomly selected packet from its generation to its successful delivery. For a stable system, the average latency per SF (including the packet on service) for Geo/PH/1 (Geo/Geo/ for $q = 1$) is obtained, following the steps in [13, Ch.5.4, Ch.5.8], as:

$$\mathbb{E}\left[W_q\right] = \begin{cases} \frac{\bar{a}a x_0}{(p_q - a)^2} + \frac{1}{p_q}, \text{for } q = 1, \\ \frac{\boldsymbol{x}_{q,1}\left(\boldsymbol{I} - \boldsymbol{R}_q\right)^{-2}\boldsymbol{e}}{a}, \text{ for } q > 1. \end{cases} \tag{13}$$

IV. Numerical Results

Monte Carlo simulations, with 10^4 iterations, are used to validate p_q by spatially averaging over the active nodes. We consider a random, yet fixed instantiation of a LoRa network, where SFs are assigned following an EIB scheme. In each simulation step, channel gains are instantiated, packets arrive following a geometric distribution, DC states are updated, and transmissions occur whenever the device buffer is nonempty, and its DC is satisfied. Unless otherwise stated, $R = 1$ km, $\eta = 3$, $\Gamma_1 = [-20, -23, -26, -29, -32, -35]$ (dB), $a = 0.02$ (packets/slot), $\lambda_L = 10^3/\text{km}^2$, $\mathcal{P} = 14$ dBm, and NF= 6 dBm.

In Fig.2a, we plot the analytical and simulated coverage probability of each SF versus different SINR thresholds. Fig.2a shows the achieved reliability level of different SFs as the capture threshold varies. The reliability is improved as the SF decreases. To maintain stable buffers, higher SF needs lower capture thresholds. A good agreement between analytical (Lines) and simulation (Markers) is shown. Moreover, Fig.2a compares the three assumptions of orthogonality as well as the case of saturated buffer. We can clearly see that ignoring traffic dynamics will lead to an aggressive interference scenario making,hence, the coverage prediction pessimistic.

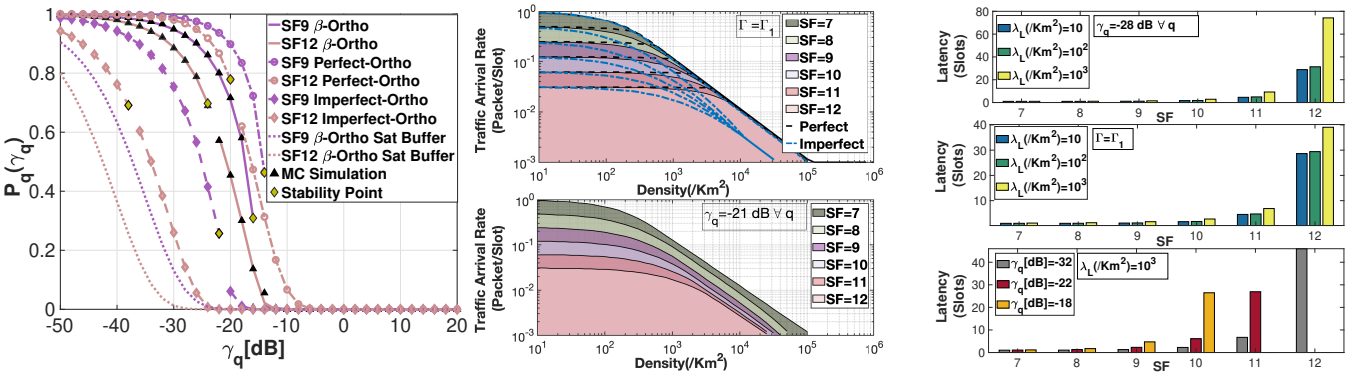
Algorithm 1: Computation of $\boldsymbol{\varphi}_q$ and $p_q \forall q$

Data: N, λ_L, Γ, a, ϵ
Initialisation: $k = 0$, $\boldsymbol{x}_{q,0}$ and $\boldsymbol{\varphi}_q$ s.t. $\boldsymbol{x}_{q,0}\boldsymbol{e} + \boldsymbol{\varphi}_q\boldsymbol{e} = 1 \forall q$;
while $\max_q\{\|\boldsymbol{\varphi}_q^{(k+1)} - \boldsymbol{\varphi}_q^{(k)}\|\} \geq \epsilon$ **do**
  for *q in Q* **do**
    1. Evaluate p_q following Eq.1 and using $\boldsymbol{\Delta}^{(k+1)}$;
    2. Construct $\boldsymbol{F}_q$ and hence $\boldsymbol{A}_q$ using p_q;
    3. Obtain $\boldsymbol{\pi}_q$ from Lemma 1 to check stability:
    if *Stable* **then**
      3.1 Calculate $\boldsymbol{R}_q^{(k+1)}$ using (9);
      3.2 Solve $\boldsymbol{x}_{q,0}$ and $\boldsymbol{x}_{q,1}$ using Theorem 2;
      3.3 Compute $\boldsymbol{\varphi}_q^{(k+1)} = \boldsymbol{x}_{q,1}\left(\boldsymbol{I} - \boldsymbol{R}_q^{(k+1)}\right)^{-1}$;
    else
      3.4 Set $\boldsymbol{x}_{q,0} = 0$ and $\boldsymbol{\varphi}_q^{(k+1)} = \boldsymbol{\pi}_q$
    end
  end
  Increment k ;
end
Return $\boldsymbol{R}_q \leftarrow \boldsymbol{R}_q^{(k+1)}$, $\boldsymbol{\varphi}_q \leftarrow \boldsymbol{\varphi}_q^{(k+1)}, \forall q$

Fig.2b represents the Pareto stability region for each SF. The result clearly shows the compromise between the scalability (number of deployed devices) and stability, i.e. going for higher density requires a decrease in arrival rate to maintain unsaturated buffers and consequently finite delay to deliver packets. Even with lower sensitivities, higher SFs can drastically hinder the traffic load. Indeed, SFs> 10 fail to accommodate sensing applications with an arrival rate higher than 0.13 packets/slot, even for densities as low as 10 nodes/km^2. The rationale behind this is that higher SFs spend a longer time in DC, leading to nodes vacating their buffers less often and consequently compromising the stability of the network.

By examining Fig.2b, an aspect worth analysing is that the maximal arrival rate stabilising the buffers for all the SFs converges to the same value, that of the lowest SF. Two main reasons come into play to explain this behaviour. First, the threshold distribution Γ_1 improves higher SFs' capture capabilities compared to $\gamma_q = -21$ dB. Second, as the density increases, the interference effect becomes more prominent; and although higher SFs transmit less often due to their longer DC, their spatio-temporal intensity (i.e. interferers/time) becomes comparable to that of lower SFs given the static EIB SF allocation scheme. A valuable conclusion drawn from the previousf analysis is that SF allocation is a determining factor for LoRa scalability. As such, SF assignment strategies that adapt to the nodes environment (i.e. channel conditions, interference severity, etc...) merit future scalability study.

Fig.2b also illustrates the impact of different orthogonality assumptions (perfect $\beta_{q_1 \neq q_2} = 0$, and imperfect $\beta_{q_1,q_2} = 1$ $\forall\{q_1, q_2\}$) on the stability frontiers. A proper orthogonality quantification would extend the scalability limits allowing a more accurate planning for massive LoRa networks. Indeed, while the different assumptions leads to the same Per-SF maximum arrival rate for densities < 20 nodes/km^2, the difference between becomes more noteworthy for $\lambda \in [10^2 - 10^4]$ nodes/km^2, a range that spans densities in practical deploy-



(a) Coverage Probability (b) Stability Frontiers (c) End-To-End Latency

Fig. 2: Coverage probability versus capture thresholds , Stability frontiers (λ_L, a) and Per-SF average latency

ment scenarios. For a traffic arrival rate of 10^{-3} packets/slot, the maximum density is 2×10^4 nodes/km^2 with imperfect-orthogonality, instead of 10^5 nodes/km^2 with β_{q_1,q_2}.

Fig.2c shows the impact of density and capture thresholds on the latency. For SFs$<$ 10, the average latency is resilient to the increase in density and remains less than 3 slots in all cases. The impact of higher density, implying more mutual interference, is more notable on SF= 12, as latency reaches 74 slots for $\lambda_L = 10^3$ nodes/km^2. A $\gamma_q \geq -18$ dB makes the buffers of devices with SF$\in \{11, 12\}$ overflow leading to unbounded delays. As such, an SF$\in \{11, 12\}$ fails to accommodate even best-effort delay-tolerant applications. Increasing γ_q from -22 dB to -18 dB makes the latency of SF= 10 jumps from 6 to 26 slots. Thus, a proper distribution of the capture thresholds among the SFs would mitigate the unfairness in terms of latency between them.

V. Conclusion

This paper proposes a joint scalability-stability analysis of LoRa. The proposed model accounts for various random factors (channel, interference, etc.), includes key LoRa parameters (imperfect-orthogonality, DC), and adopts a more accurate quantification of imperfect-orthogonality. The stability frontiers of each SF are analysed. The results show the compromise between the network stability and scalability illustrated by the limits of nodes' density and per-device traffic rate that a LoRa SF can accommodate. They also highlight the importance of SF adaptation and capture thresholds distribution toward a massive LoRa network and improved latency.

References

[1] Y. Bouazizi, F. Benkhelifa, and J. McCann, "Spatiotemporal modelling of multi-gateway LoRa networks with imperfect SF orthogonality," in *IEEE Global Communications Conference (Globecom'2020)*, 2020, pp. 1–7.

[2] O. Georgiou and U. Raza, "Low power wide area network analysis: Can LoRa scale?" *IEEE Wireless Communications Letters*, vol. 6, no. 2, pp. 162–165, April 2017.

[3] J. Lim and Y. Han, "Spreading factor allocation for massive connectivity in LoRa systems," *IEEE Communications Letters*, vol. 22, no. 4, pp. 800–803, April 2018.

[4] L. Beltramelli, A. Mahmood, M. Gidlund, P. Österberg, and U. Jennehag, "Interference modelling in a multi-cell LoRa system," in *14th International Conference on Wireless and Mobile Computing, Networking and Communications (WiMob'2018)*, October 2018, pp. 1–8.

[5] R. B. Sørensen, D. M. Kim, J. J. Nielsen, and P. Popovski, "Analysis of latency and MAC-layer performance for class a LoRaWAN," *IEEE Wireless Communications Letters*, vol. 6, no. 5, pp. 566–569, 2017.

[6] R. M. Sandoval, A.-J. Garcia-Sanchez, and J. Garcia-Haro, "Performance optimization of LoRa nodes for the future smart city/industry," *EURASIP journal on wireless communications and networking*, vol. 2019, no. 1, pp. 1–13, 2019.

[7] D. Croce, M. Gucciardo, S. Mangione, G. Santaromita, and I. Tinnirello, "Impact of LoRa imperfect orthogonality: Analysis of link-level performance," *IEEE Communications Letters*, vol. 22, no. 4, pp. 796–799, April 2018.

[8] F. Benkhelifa, Y. Bouazizi, and J. A. McCann, "How orthogonal is lora modulation?" *IEEE Internet of Things Journal*, pp. 1–1, 2022.

[9] A. Waret, M. Kaneko, A. Guitton, and N. El Rachkidy, "Lora throughput analysis with imperfect spreading factor orthogonality," *IEEE Wireless Communications Letters*, vol. 8, no. 2, pp. 408–411, 2019.

[10] R. B. Sorensen, N. Razmi, J. J. Nielsen, and P. Popovski, "Analysis of lorawan uplink with multiple demodulating paths and capture effect," in *ICC 2019 - 2019 IEEE International Conference on Communications (ICC)*, 2019, pp. 1–6.

[11] F. Benkhelifa, Z. Qin, and J. A. McCann, "User fairness in energy harvesting-based LoRa networks with imperfect SF orthogonality," *IEEE Transactions on Communications*, pp. 1–1, 2021.

[12] Alan Jeffrey, Daniel Zwillinger, I. Gradshteyn, and I. Ryzhik, Eds., *"Table of Integrals, Series, and Products (Seventh Edition)"*, 7th ed. Boston: Academic Press, 2007, pp. 1063 – 1068.

[13] A. S. Alfa, *Applied Discrete-Time Queues*, 2nd ed. Springer Publishing Company, Incorporated, 2015.

[14] I. Krikidis, T. Charalambous, and J. S. Thompson, "Buffer-aided relay selection for cooperative diversity systems without delay constraints," *IEEE Transactions on Wireless Communications*, vol. 11, no. 5, pp. 1957–1967, 2012.

Appendix A
Laplace Transform of $\mathcal{I}$

During a generic time slot, we have:

$$\begin{aligned}\mathcal{L}_{\mathcal{I}}\{s\} &= \mathbb{E}_{\phi,G}\left[e^{-s\sum_{q\in\mathcal{Q}}\sum_{x_j\in\tilde{\Phi}_{L,q}\setminus x_0}\mathcal{P}\alpha\|^{-\eta}\beta_{q_0,q}g_j\|x_j}\right]\\ &\overset{(a)}{=} \mathbb{E}_{\phi}\left[\prod_{q=1}^{N}\prod_{x_j\in\tilde{\Phi}_{L,q}\setminus x_0}\frac{1}{1+s\mathcal{P}\beta_{q_0,q}\alpha\|x_j\|^{-\eta}}\right]\\ &\overset{(b)}{\approx} \prod_{q=1}^{N}\left[\mathbb{E}_{x}\prod_{x_j\in\tilde{\Phi}_{L,q}\setminus x_0}\frac{1}{1+s\mathcal{P}\beta_{q_0,q}\alpha\|x_j\|^{-\eta}}\right]\\ &\overset{(c)}{=} \prod_{q=1}^{N} e^{-2\pi\delta_q\lambda_L\int_{d_{q-1}}^{d_q}\left(\frac{s\mathcal{P}\beta_{q_0,q}\alpha r^{-\eta}}{1+s\mathcal{P}\beta_{q_0,q}\alpha r^{-\eta}}\right)rdr},\end{aligned} \tag{14}$$

where (a) follows the moment generating function (MGF) of the unit-mean exponential distribution with independence of channel gains from the space and time (b) is an approximation obtained using Fortuin–Kasteleyn–Ginibre (FKG) inequality, (c) follows the probability generating function (PGFL) of $\tilde{\Phi}_L$.